PROPOSAL FOR TOPIC APPROVAL TO BE SUBMITTED TO THE UNIVERSITY OF MUMBAI FOR THE DEGREE OF Ph.D. (SCIENCE AND TECHNOLOGY) IN THE SUBJECT OF COMPUTER ENGINEERING

Title of Topic	:	Deepfake Detection System for Identification and Evidence Integrity in Digital Forensics using Mumtimodal Artificial Intelligence Driven Approaches
Name of the Student	:	Mrs. Manisha Prabhakar Sonawane
Basic Qualification	:	M.E.(Computer Engineering) B.E.(Computer Engineering)
In the Case of Ph.D. student Year of Passing of PET/NET/SET/GATE (Teaching Experience)	:	PET - 2024
Name of the Research Supervisor/ Guide	:	Dr. Uttara D. Gogate
Date of Superannuation / Retirement	:	16/03/2036
Name of the Research Center	:	Shah and Anchor Kutchhi Engineering College
Signature of Student	:	
Signatue of the Research Supervisor	:	

Shah & Anchor Kutchhi Engg. College * Autonomous

Dr. BHAVESH PATEL
Principal
Shah & Anchor Kutchhi Engineering College
Autonomous Institute
Mahavir Education Trust Chowk,
W. T. Patil Marg, Next to D-Mart,
Chembur - 400088

Scanned with OKEN Scanner

Sr. no.	Content	Page no.
1	Introduction	1
2	Literature Review	1
3	Aim & Objectives	4
4	Materials & Methods	5
5	Expected Outcome	6
6	References	7

Scanned with OKEN Scanner

Statement by the Candidate

I wish to state that the work embodied in this synopsis titled " Deepfake Detection System for Identification and Evidence Integrity in Digital Forensics using Mumtimodal Artificial Intelligence Driven Approaches" forms my own contribution to the work carried out under the guidance of Dr. Uttara D. Gogate at the Shah and Anchor Kutchhi Engineering College. I declare that this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

(Candidate Signature)

Name: Mrs. Manisha Prabhakar Sonawane

Roll Number: 124PHCM1008

Scanned with OKEN Scanner

Introduction

Deepfake detection techniques encompass a diverse range of methodologies in distinguishing authentic media from AI-generated manipulations. These methods leverage advanced neural network architectures, frequency domain analysis, and ensemble strategies, enabling robust identification of both image and video deepfakes across varied datasets and manipulation types.[1] Deepfakes have also been used to fabricate satellite imagery by introducing non-existent landscape features for malicious purposes. Beyond this, deepfake technology finds intriguing applications in areas such as video compositing and portrait transfiguration — for example, in identity protection, where real faces in photographs can be replaced with those from a curated collection of stock images.

Meanwhile, cyber-attackers continually seek to compromise identification and authentication systems, often employing methods beyond deepfakes to gain unauthorized access. Detecting deepfake media using forensic approaches remains a significant challenge, as attackers quickly adapt by integrating the latest detection methods into new generations of deepfake creation techniques[2].

The explosive growth of the Internet and social media — coupled with billions of publicly available images — has led to a considerable erosion of trust among users. Deepfakes now pose a major threat not only to society but also to the credibility of digital evidence presented in courts. Therefore, it is critically important to develop and adopt state-of-the-art detection techniques, particularly for criminal investigations[1][2].

Broadly, the techniques used to manipulate multimedia content can be grouped into four main categories: copy–move, splicing, deepfake, and resampling[3]. Copy–move involves duplicating and repositioning parts of an image; splicing merges regions from different images into a composite; and resampling modifies scale and orientation of image elements. All these methods aim to deceive viewers by presenting more visual elements than were originally captured.

Deepfake manipulation, however, harnesses advanced machine learning (ML) techniques to create highly convincing fake content. It can be viewed as an evolved form of splicing — where a person's face, voice, or actions in videos or images are replaced with those of a fabricated target[3]. These manipulations are often linked to a broad spectrum of cybercrime activities; while distributing deepfakes is relatively simple, correcting misinformation and preventing their spread remains far more difficult. Deepfake detection is a rapidly evolving field leveraging deep learning architectures, ensemble strategies, frequency analysis, and hierarchical classification to combat increasingly sophisticated media manipulations. The integration of diverse datasets, advanced preprocessing, and ethical considerations underpins the development of reliable systems essential for safeguarding digital integrity in modern society.[3] The proliferation of deepfakes—hyper-realistic fake images, audio, and videos generated by deep generative models such as GANs and diffusion models—has presented significant challenges for digital media integrity, legal proceedings, and public trust.

Literature Review

Papers [4] to [13] introduce a diverse range of deepfake detection methods, each targeting different aspects of video and image manipulation. Fuzzy-DFD [4] uses a multi-stream ResNet50 architecture with a Fuzzy Inference System to robustly fuse spatial, temporal, and structural features. Paper [5] presents an unsupervised method based on PRNU and noiseprint to detect anomalies in camera fingerprints without labeled data. Paper [6] combines 1D and 2D CNNs with attention for deepfake detection, outperforming traditional ML models. The method in [7] enhances robustness across domains using perturbation alignment via PIMW and an MLP projector. Paper [8] boosts accuracy

Scanned with OKEN Scanner

using Local Surface Geometry features extracted by UprightNet and classified with deep networks like ResNet. Paper [9] focuses on anomaly detection from real video training only, using facial-region comparisons. The encoder-decoder model in [10] disentangles forgery from non-forgery features to improve generalization across datasets. Paper [11] enhances detection by applying Discrete Wavelet Transform preprocessing to emphasize frequency-domain features. UNTAG [12] proposes a type-agnostic, unsupervised framework that learns solely from real data using a self-supervised task and GMM classifier. Finally, paper [13] introduces a graph-based dual-branch model incorporating RGB and color distortion features with modular components to improve generalization across datasets.

Papers [14] to [24] present advanced deepfake and synthetic image detection methods leveraging deep learning, frequency analysis, and interpretability techniques. Paper [14] combines FaceNet and ResNet50 to fuse time- and frequency-domain features for improved fake face detection. Paper [15] mitigates identity bias with an ID-unaware model that emphasizes artifact-based cues. A VGG19-based hybrid model in [16] integrates spatial and wavelet-based frequency features to boost cross-domain performance. Paper [17] studies attack transferability between CNN and CLIP detectors, suggesting combined use enhances robustness. Paper [18] uses spectral domain analysis to amplify abnormal frequency patterns, achieving high interpretability and resilience. Paper [19] identifies DCT-based BetaAC coefficients as key markers for classifying real vs. GAN/DM-generated images, maintaining robustness under compression. FILTER [20] targets multi-face forgeries with novel components to enhance efficiency and accuracy. Paper [21] exposes cybersecurity risks in ISAR systems via GAN-generated radar image forgeries. Paper [22] introduces DDL, an interpretability framework using feature attribution and facial parsing for transparent deepfake detection. DeePhyNet [23] combines spatial-temporal features to detect deepfakes and trace their generative origin, supported by the DeePhyV2 dataset. Lastly, paper [24] presents the Semantic Token Transformer (STT), integrating facial semantics into transformers to significantly enhance detection accuracy.

Papers [25] to [31] explore various deepfake detection techniques emphasizing spatial-temporal analysis, model efficiency, and robustness. Papers [25] and [26] both propose similar approaches that integrate ResNeXt-50 for spatial feature extraction with LSTM for capturing temporal dynamics, achieving high accuracy on curated datasets. Paper [27] reviews face anti-spoofing methods, advocating for the integration of manipulation detection with liveness verification to enhance facial recognition security. Paper [28] introduces a lightweight CNN that outperforms state-of-the-art models on the FaceForensics dataset, with ablation studies highlighting key design factors. Paper [29] presents a contrastive learning approach comparing whole-face and center-face features, improving generalization and robustness, especially on common manipulation types and compressed images. Paper [30] investigates video forgery detection, including deepfakes and copy-paste attacks, under varying compression levels using computer vision and machine learning techniques. Lastly, paper [31] evaluates the performance of CNNs like EfficientNet, XCeptionNet, and ResNet, finding EfficientNet most effective, though enhancements like Noisy Student and Face Cutout provided limited benefit.

Papers [32] to [37] present advanced deepfake detection methods emphasizing spatial-temporal modeling, multimodal feature integration, and robustness. Paper [32] leverages facial landmark calibration with Transformers and bi-directional LSTMs to extract geometric and temporal features, enhancing compression robustness. Paper [33] introduces DeepStream-X, a two-stream framework combining high-frequency features and motion information using Xception with CBAM, achieving state-of-the-art results on Celeb-DF v2. Paper [34] evaluates 12,500 videos using CNNs with and without GRUs, finding Xception and DenseNet121+GRU most effective. Addressing cross-forgery generalization, paper [35] fuses RGB and optical flow data using a CNN-ViT hybrid, significantly outperforming single-stream models. Paper [36] proposes ISTVT, a transformer-based model using decomposed spatial-temporal attention and a self-subtract mechanism to enhance interpretability and generaliza-

Scanned with OKEN Scanner

tion. Lastly, paper [37] integrates ResNet50, LSTM, and optical flow to capture both visual and motion cues, improving deepfake video detection accuracy.

Papers [38] to [47] introduce innovative deepfake detection methods focused on anomaly detection, stylistic inconsistencies, geometric features, and robust training strategies. DiffFake [38] uses pseudo-deepfakes with EfficientNet-b4 and a GMM anomaly module to improve cross-dataset generalization. Paper [39] exploits latent style inconsistencies using StyleGRU and Style Attention Modules for superior temporal-stylistic fusion. Paper [40] reframes detection as time-series anomaly analysis using facial landmark trajectories and region-based autoencoders. FAMM [41] detects deepfakes via facial muscle movement from landmark geometry, excelling under compression. DualFlat optimization [42] improves model generalization by focusing on data distribution flatness. Paper [43] applies 3D decomposition and neural architecture search (NAS) for high-accuracy forgery detection, especially for unseen manipulations. Paper [44] uses adversarial training with pixel-level blurring and generator-based examples to enhance robustness to novel forgeries. Paper [45] applies ResNet-50 to detect fake news via image analysis, achieving high accuracy but noting computational challenges. TruFor [46] fuses RGB data with a noise-sensitive fingerprint (Noiseprint++) for effective detection and localization, even under compression. Finally, MFM-LCFI [47] boosts detection of subtle clues in low-quality videos by combining multi-feature enhancement with dual-frequency decomposition.

Papers [48] to [50] propose innovative strategies for deepfake detection focusing on watermarking, group-level analysis, and hybrid models. Paper [48] presents PWPH, a proactive framework embedding perceptual watermarks in non-face regions while maintaining image quality, achieving superior watermark recovery and detection performance. Paper [49] explores pooled deepfake video analysis using PRNU autocorrelation, revealing that group-level fingerprint patterns can enhance detection beyond typical re-encoding cues. Paper [50] introduces a hybrid method combining a Spatiotemporal Convolutional Network with PRNU analysis, effectively integrating deep learning and sensor noise for high detection accuracy and speed across multiple datasets.

Scanned with OKEN Scanner

Aim & Objectives

To develop and evaluate advanced, robust deepfake detection methodologies that leverage multi-modal digital forensic techniques, enhancing the accuracy and reliability of identifying manipulated media across diverse real-world scenarios, including cross-modality detection and scalability

1. To identify issues in existing deepfake detection systems and design efficacious detection technique.
2. To examine digital media for deepfake manipulation.
3. To analyse proposed system to find an optimal solution for deepfake detection.
4. To evaluate the performance of the proposed system on real dataset.

Scanned with OKEN Scanner

Materials & Methods

This work proposes a novel approach to the detection of fakes in digital forensics for image, video, and audio.Acquisition of input is done form dataset.If the input is Image then the pre-prossessing is done and feature extraction, face detection is done in Feature Extraction phase. Deep learning and frequency domain analysis methods are used to determine whether the image is fake or real. If the input is video, then frame extraction is performed. Frames are analyzed using deep learning, frequency domain analysis, and biological signal analysis to detect frame changes and classified as fake or real. If the input is audio, then the audio is preprossessed and given to artifact detection method and classified as real or fake.

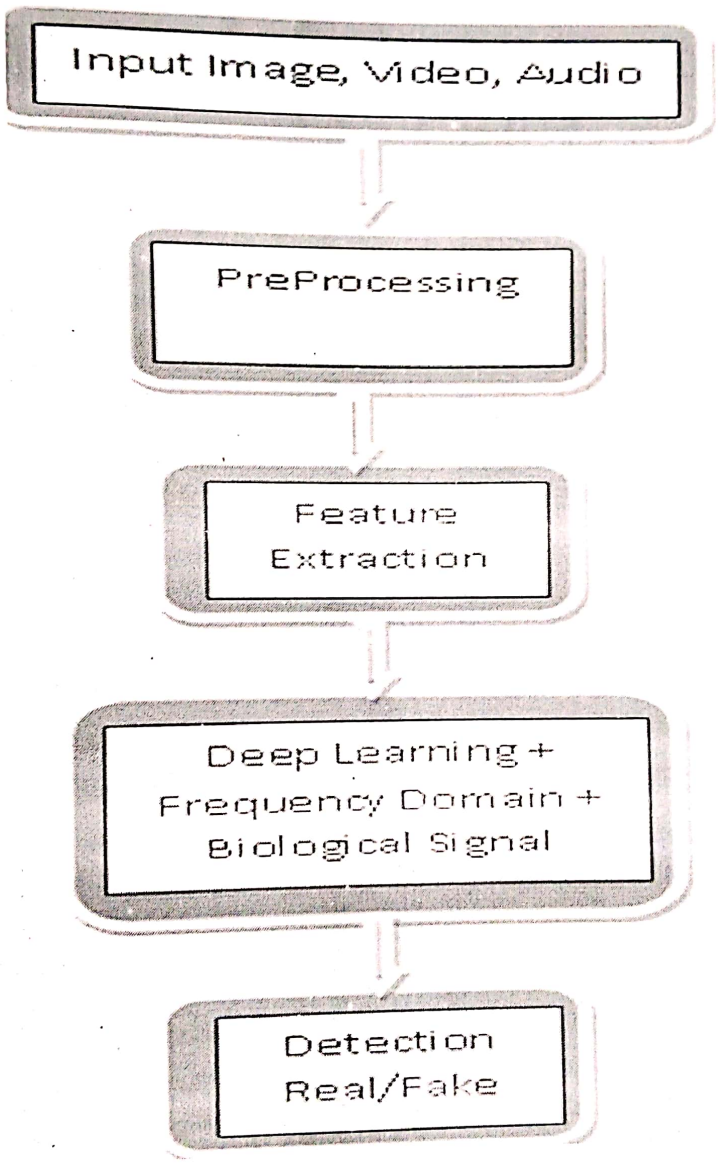

Fig 1. Deepfake Detection Framework

Scanned with OKEN Scanner

Expected Outcomes

1. Gain a detailed understanding of the limitations and vulnerabilities in current deepfake detection methods, leading to the design of a more effective and resilient detection technique.
2. A systematic approach for inspecting and analyzing digital media to detect potential manipulations.
3. Offering a balance between accuracy, speed, and scalability.
4. Insights into false positive/negative cases.

Scanned with OKEN Scanner

References

[1] N. Ur Rehman Ahmed, A. Badshah, H. Adeel, A. Tajammul, A. Daud and T. Alsahfi, "Visual Deepfake Detection: Review of Techniques, Tools, Limitations, and Future Prospects," in IEEE Access, vol. 13, pp. 1923-1961, 2025.

[2] R. Ranout and C. Kumar, "Unmasking the Illusions: A Comprehensive Study on Deepfake Videos and Images," 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-7.

[3] F. Mira, "Deep Learning Technique for Recognition of Deep Fake Videos," 2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET), London, United Kingdom, 2023, pp. 1-4.

[4] H. She, Y. Hu, B. Liu, J. Li and C. -T. Li, "Using Graph Neural Networks to Improve Generalization Capability of the Models for Deepfake Detection," in IEEE Transactions on Information Forensics and Security, vol. 19, pp. 8414-8427, 2024.

[5] M. Wang, P. Fan and T. Yang, "Fake face detection based on deep learning and frequency domain processing," 2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, 2023, pp. 77-82.

[6] B. Subburaj and R. Ragavendra, "Deepfake Detection Using Spatio-Temporal-Structural Anomaly Learning and Fuzzy System-Based Decision Fusion," in IEEE Access, vol. 13, pp. 82747-82758, 2025.

[7] Li Zhang, Tong Qiao, Ming Xu, Ning Zheng, and Shichuang Xie. 2023. Unsupervised Learning-Based Framework for Deepfake Video Detection. Trans. Multi. 25 (2023), 4785–4799.

[8] R. B. P. and M. S. Nair, "Deepfake Detection using Multi-path CNN and Convolutional Attention Mechanism," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-6.

[9] L. Lu, Y. Wang, L. Zhang and Y. Guo, "Robust Deepfake Detection via Perturbation Domain Alignment," ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5.

[10] M. Sivabalamurugan and T. R. Swapna, "Deepfake detection and classification using local surface geometrical features," 2024 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI), Prayagraj, India, 2024, pp. 1-6.

[11] Y. Wang and G. Liao, "Deepfake Video Detection Based on Image Source Anomaly," 2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA), Shenyang, China, 2024, pp. 397-401.

[12] J. Hu, S. Wang and X. Li, "Improving the Generalization Ability of Deepfake Detection via Disentangled Representation Learning," 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 2021, pp. 3577-3581.

[13] S. J. Burroughs, B. Gokaraju, K. Roy and L. Khoa, "DeepFakes Detection in Videos using Feature Engineering Techniques in Deep Learning Convolution Neural Network Frameworks," 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington DC, DC, USA, 2020.

Scanned with OKEN Scanner

[14] N. Mejri, E. Ghorbel and D. Aouada, "UNTAG: Learning Generic Features for Unsupervised Type-Agnostic Deepfake Detection," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5.

[15] S. Dong, J. Wang, R. Ji, J. Liang, H. Fan and Z. Ge, "Implicit Identity Leakage: The Stumbling Block to Improving Deepfake Detection Generalization," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 3994-4004.

[16] S. Sadhya and X. Qi, "Deepfake Detection with Wavelet-Integrated Convolutional Networks," 2024 IEEE International Conference on Big Data (BigData), Washington, DC, USA, 2024, pp. 4540-4547.

[17] V. De Rosa, F. Guillaro, G. Poggi, D. Cozzolino and L. Verdoliva, "Exploring the Adversarial Robustness of CLIP for AI-generated Image Detection," 2024 IEEE International Workshop on Information Forensics and Security (WIFS), Rome, Italy, 2024, pp. 1-6.

[18] Y. Li et al., "MaskSim: Detection of synthetic images by masked spectrum similarity analysis." 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA, 2024, pp. 3855-3865.

[19] O. Pontorno, L. Guarnera and S. Battiato, "On the Exploitation of DCT-Traces in the Generative-AI Domain," 2024 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 2024, pp. 3806-3812.

[20] C. Lin et al., "Exploiting Facial Relationships and Feature Aggregation for Multi-Face Forgery Detection," in IEEE Transactions on Information Forensics and Security, vol. 19, pp. 8832-8844. 2024.

[21] A. H. Oveis, G. Meucci, F. Mancuso, F. Berizzi and A. Cantelli-Forti, "Generative AI Threats to Maritime Navigation Using Deceptive ISAR Images," in IEEE Access, vol. 12, pp. 173800-173809. 2024.

[22] Z. Sun, N. Ruan and J. Li, "DDL: Effective and Comprehensible Interpretation Framework for Diverse Deepfake Detectors," in IEEE Transactions on Information Forensics and Security, vol. 20, pp. 3601-3615, 2025.

[23] K. Thakral, H. Agarwal, K. Narayan, S. Mittal, M. Vatsa and R. Singh, "DeePhyNet: Toward Detecting Phylogeny in Deepfakes," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 7, no. 1, pp. 132-145.

[24] Chunlei Peng, Xiaoyi Luo, Decheng Liu, Nannan Wang, Ruimin Hu, and Xinbo Gao. 2025. Semantic Token Transformer for Face Forgery Detection. Trans. Info. For. Sec. 20 (2025), 4904–4914.

[25] A. Barbadekar, S. Sole and A. Shekhavat, "Enhancing Social Media Security: LSTM-Based Deep Fake Video Detection," 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-6.

[26] Sun, R., Zhao, Z., Shen, L., Zeng, Z., Li, Y., Veeravalli, B., and Yang, X." An Efficient Deep Video Model for Deepfake Detection". 2023 IEEE International Conference on Image Processing (ICIP) Poster Session. IEEE.

Scanned with OKEN Scanner

[27] R. B. Hadiprakoso, H. Setiawan and Girinoto, "Face Anti-Spoofing Using CNN Classifier and Face liveness Detection," 2020 3rd International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2020, pp. 143-147.

[28] N. Sambhu and S. Canavan, "Detecting Forged Facial Videos Using Convolutional Neural Networks," 2023 IEEE International Symposium on Smart Electronic Systems (iSES), Ahmedabad, India, 2023, pp. 260-264.

[29] Z. Ma, P. Xu, X. Mei and J. Shen, "A Novel Facial Manipulation Detection Method Based on Contrastive Learning," 2022 IEEE 5th International Conference on Electronics Technology (ICET), Chengdu, China, 2022, pp. 1163-1167.

[30] A. V and P. T. Joy, "Deepfake Detection Using XceptionNet," 2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE), Kerala, India, 2023, pp. 1-5.

[31] K. Smelyakov, Y. Kitsenko and A. Chupryna, "Deepfake Detection Models Based on Machine Learning Technologies," 2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), Vilnius, Lithuania, 2024, pp. 1-6.

[32] Q. Fu, Z. Chen, L. Li and X. Zheng, "Deepfakes Video Detection Based on Landmark Calibration and Neural Networks," 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), Shenzhen, China, 2024, pp. 997-1000.

[33] N. M. Emara and M. N. Elagamy, "DeepStream-X: A Two-Stream Deepfake Detection Framework using Spatiotemporal and Frequency Features," 2024 34th International Conference on Computer Theory and Applications (ICCTA), Alexandria, Egypt, 2024, pp. 290-296.

[34] K. R. Talwalkar, K. Navale, A. Ashok and A. Khade, "Design and Analytical Performance of a Hybrid Model Using Deep Learning Techniques for the Detection of Deepfake Videos," 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), Nagpur, India, 2024, pp. 1-7.

[35] H. Teng and C. -Y. Lin, "Dynamic and Static Features Extraction for Deepfake Detection," 2024 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), Taichung, Taiwan, 2024, pp. 123-124.

[36] C. Zhao, C. Wang, G. Hu, H. Chen, C. Liu and J. Tang, "ISTVT: Interpretable Spatial-Temporal Video Transformer for Deepfake Detection," in IEEE Transactions on Information Forensics and Security, vol. 18, pp. 1335-1348, 2023.

[37] E. M. Sathwik Reddy, A. Pavan Kumar and P. Swetha, "Deepfake video detection using CNN and RNN with OPTICAL FLOW features," 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2024, pp. 1-7.

[38] S. Stamnas and V. Sanchez, "DiffFake: Exposing Deepfakes Using Differential Anomaly Detection," 2025 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Tucson, AZ, USA, 2025, pp. 647-657.

[39] Choi, J., Kim, T., Jeong, Y., Baek, S., and Choi, J. (2024, June). Exploiting Style Latent Flows for Generalizing Deepfake Video Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1133–1143). IEEE.

Scanned with OKEN Scanner